# How far can you go with IX Route Servers only?

Ben Cartwright-Cox
BGP.Tools / Port 179 LTD

DKNOG / 2026

9933dc1dafaf3a99bb5bddd73b039965

# As I am sure you know, IX's (almost always) have services

- AS112
  - Helps soak up some DNS queries that should not have escaped a DNS recursor
- NTP/Time
  - Useful if you need to provide your router with local time
- Route Collectors
  - (sort of?) Useful if you want to double check a members policy
  - Most IX's route collectors (if they have them) are in some state of disrepair

9933dc1dafaf3a99bb5bddd73b039965

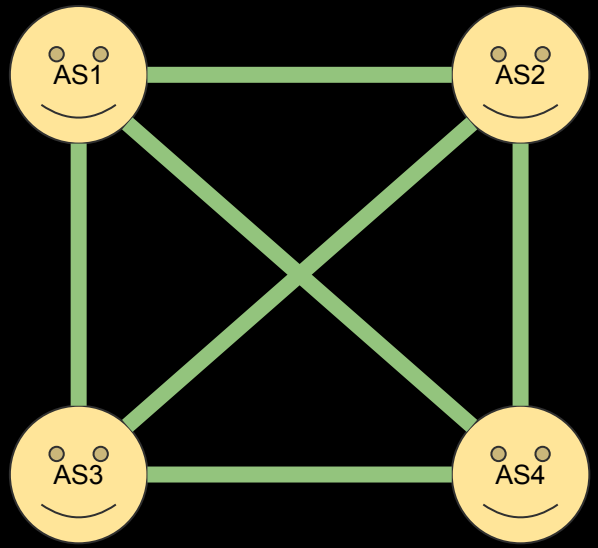
# As I am sure you know, IX's (almost always) have services

- AS112
  - Helps soak up some DNS queries that should not have escaped a DNS recursor
- NTP/Time
  - Useful if you need to provide your router with local time
- Route Collectors
  - (sort of?) Useful if you want to double check a members policy
  - Most IX's route collectors (if they have them) are in some state of disrepair
- Route Servers
  - The major service that on the IX LAN!

9933dc1dafaf3a99bb5bddd73b039965

# IX Route servers

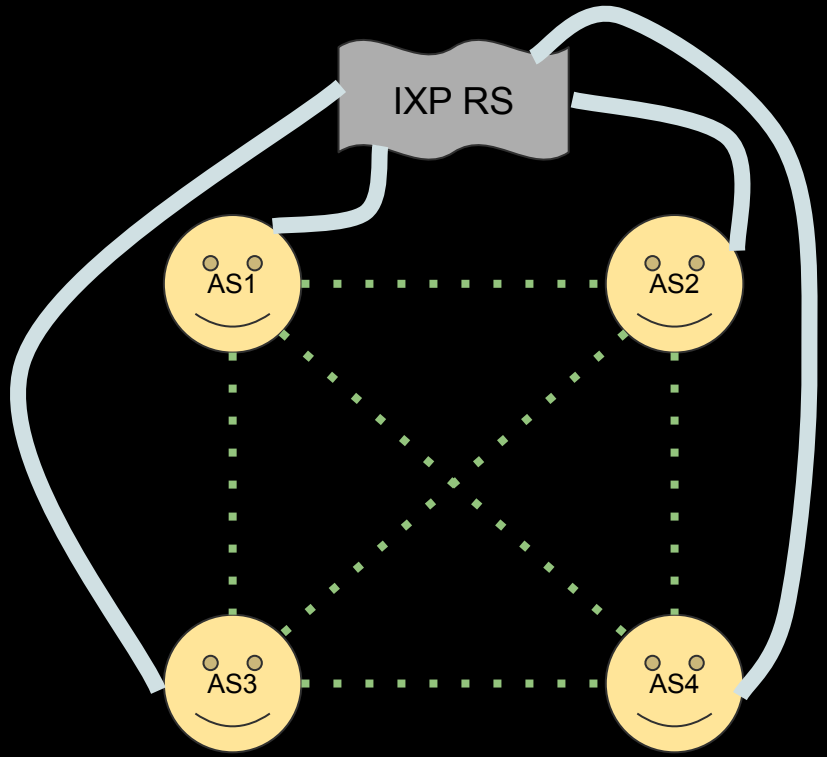
- Attempts to solve the problem of peers needing to create BGP sessions with nearly every member of a exchange they want to exchange traffic with
- This is bad because networks are lazy/busy, and may not setup sessions when asked




9933dc1dafaf3a99bb5bddd73b039965

# IX Route Servers (RS)

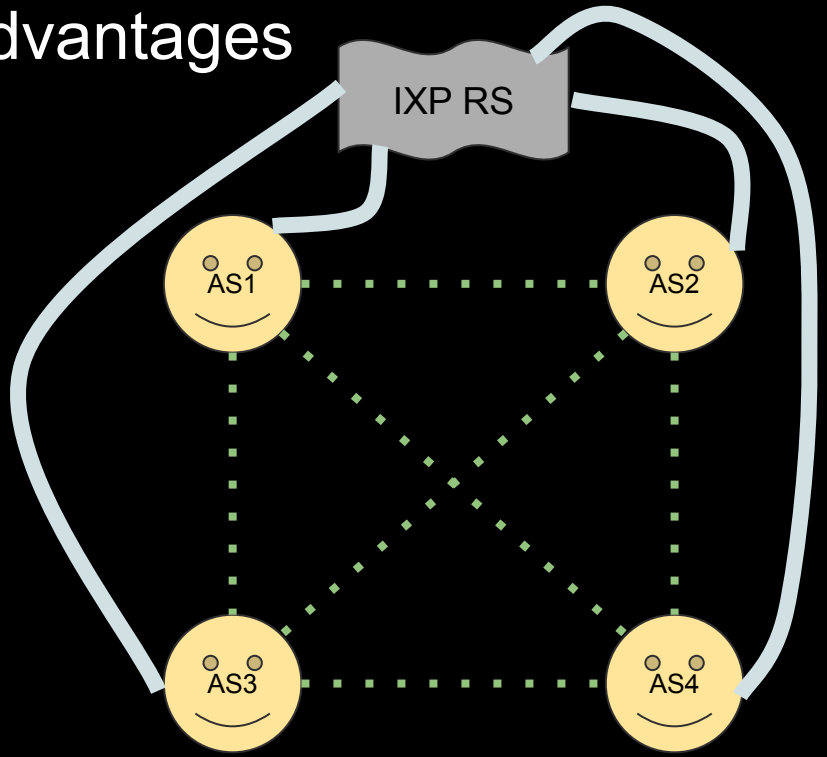
- Instead, everyone peers with the RS, and it distributes the routes sent to it by other members to everyone.
- Since everyone is on the same subnet/layer 2. The BGP next hop is not changed.
- The RS can control terabits of traffic with a 100mbps port.




9933dc1dafaf3a99bb5bddd73b039965

# IX Route Servers (RS) Bonus Advantages

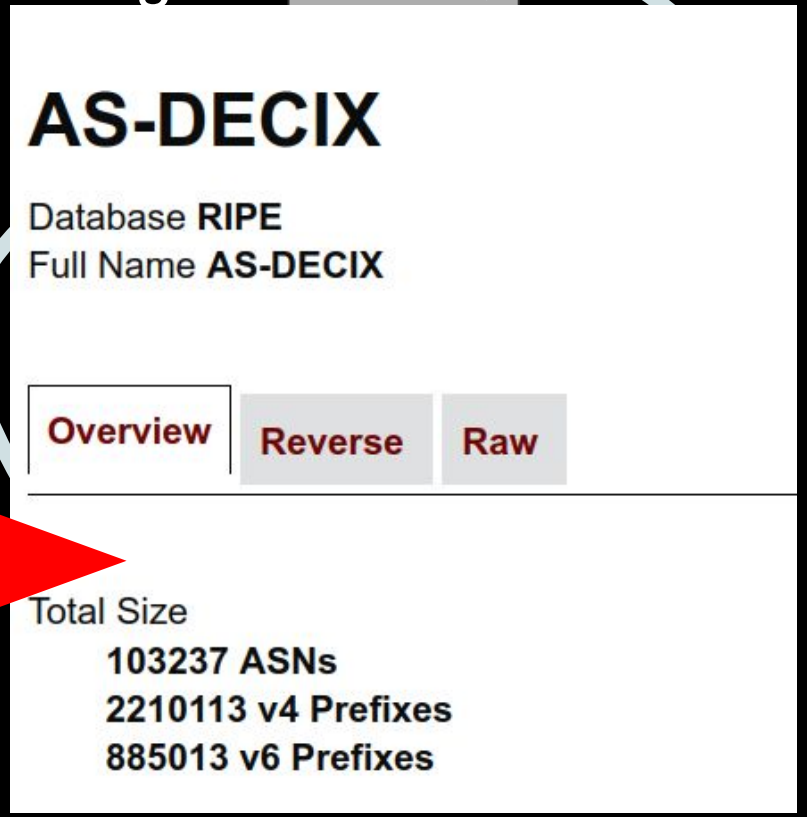
- Your average person (despite what they say publicly) does not have a good/secure BGP peer configuration
- Modern RS are ***far safer*** to peer on than bi-lat peering, due to better IRR/RPKI/Sanity/PeerLock automations
- Yes you may have the magic config, but most of the IX is importing almost anything they are sent




9933dc1dafaf3a99bb5bddd73b039965

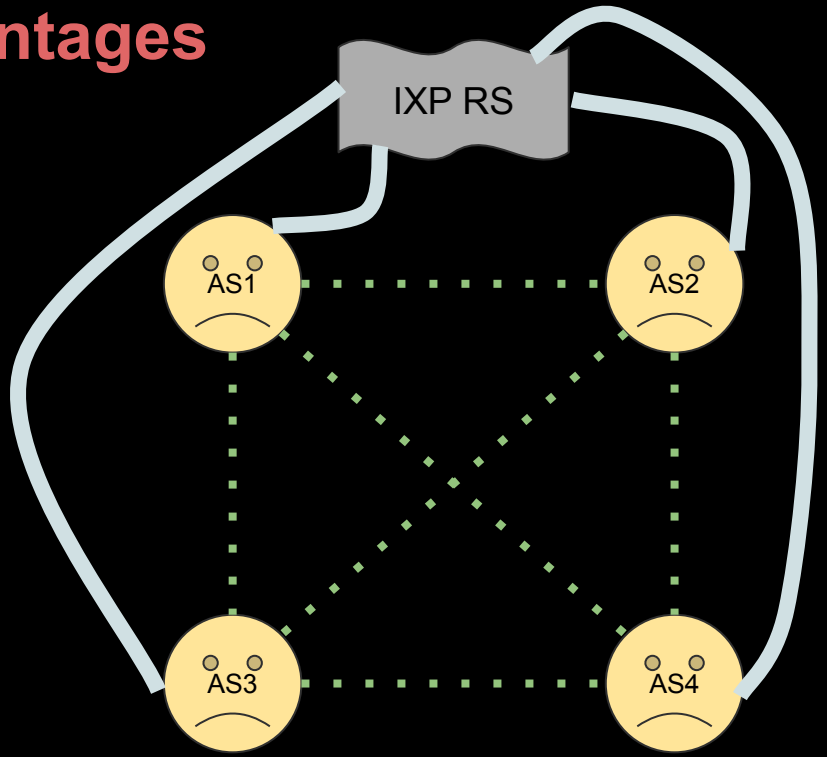
# IX Route Servers (RS) Bonus Advantages

- Your average person (despite what they say publicly) does not have a good/secure BGP peer configuration
- Modern RS are ***far safer*** to peer on than bi-lat peering, due to better IRR/RPKI/Sanity/PeerLock automations
- Yes you may have the magic config, but most of the IX is importing almost anything they are sent



# IX Route Servers (RS) **Disadvantages**

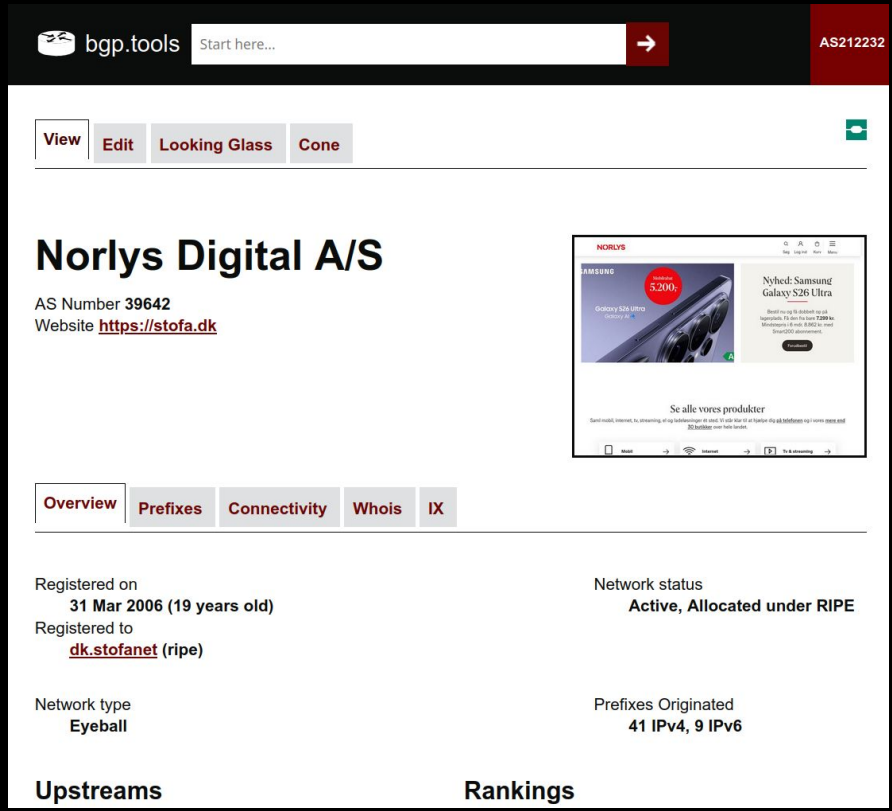
- ***Some*** networks import/export everything to the Route Server
- ***Some*** networks just import from the route server, but don't send their routes to them
    - There are some good reasons to do this, some networks are very sensitive poor routing and Route Servers are considered high risk
- ***Others*** will export/send routes, but not import anything




9933dc1dafaf3a99bb5bddd73b039965

# Quick tl;dr of what bgp.tools is

- General internet routing information site
- Runs its own BGP Route Collector that you should feed!
- Has a large IX presence for route collection
- Can also offer rapid BGP (and other BGP related things) network monitoring

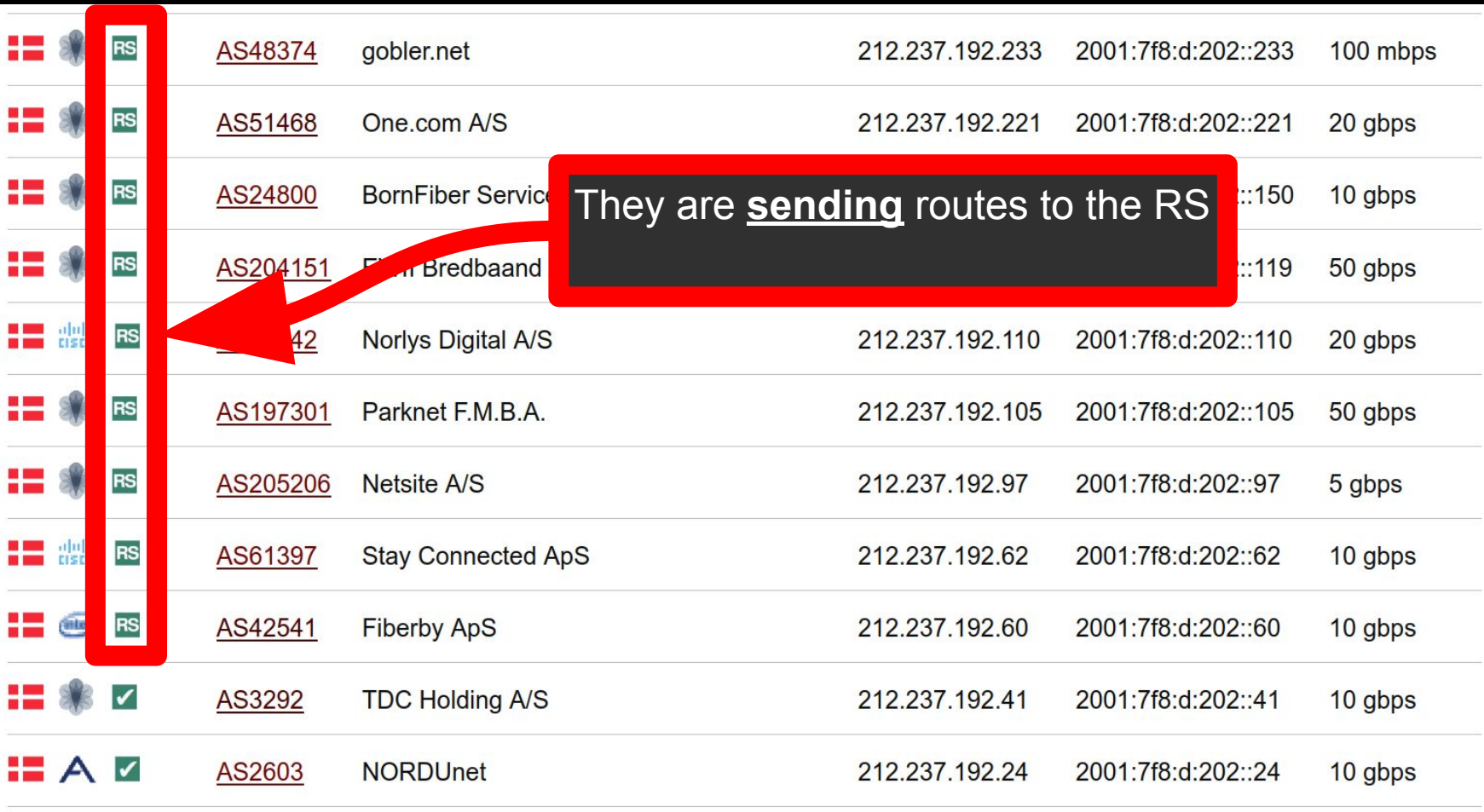



9933dc1dafaf3a99bb5bddd73b039965

| | | | | | |
|---|---|---|---|---|---|
| RS | AS48374 | gobler.net | 212.237.192.233 | 2001:7f8:d:202::233 | 100 mbps |
| RS | AS51468 | One.com A/S | 212.237.192.221 | 2001:7f8:d:202::221 | 20 gbps |
| RS | AS24800 | BornFiber Service Provider ApS | 212.237.192.150 | 2001:7f8:d:202::150 | 10 gbps |
| RS | AS204151 | EWII Bredbaand A/S | 212.237.192.119 | 2001:7f8:d:202::119 | 50 gbps |
| RS | AS39642 | Norlys Digital A/S | 212.237.192.110 | 2001:7f8:d:202::110 | 20 gbps |
| RS | AS197301 | Parknet F.M.B.A. | 212.237.192.105 | 2001:7f8:d:202::105 | 50 gbps |
| RS | AS205206 | Netsite A/S | 212.237.192.97 | 2001:7f8:d:202::97 | 5 gbps |
| RS | AS61397 | Stay Connected ApS | 212.237.192.62 | 2001:7f8:d:202::62 | 10 gbps |
| RS | AS42541 | Fiberby ApS | 212.237.192.60 | 2001:7f8:d:202::60 | 10 gbps |
| ✓ | AS3292 | TDC Holding A/S | 212.237.192.41 | 2001:7f8:d:202::41 | 10 gbps |
| ✓ | AS2603 | NORDUnet | 212.237.192.24 | 2001:7f8:d:202::24 | 10 gbps |

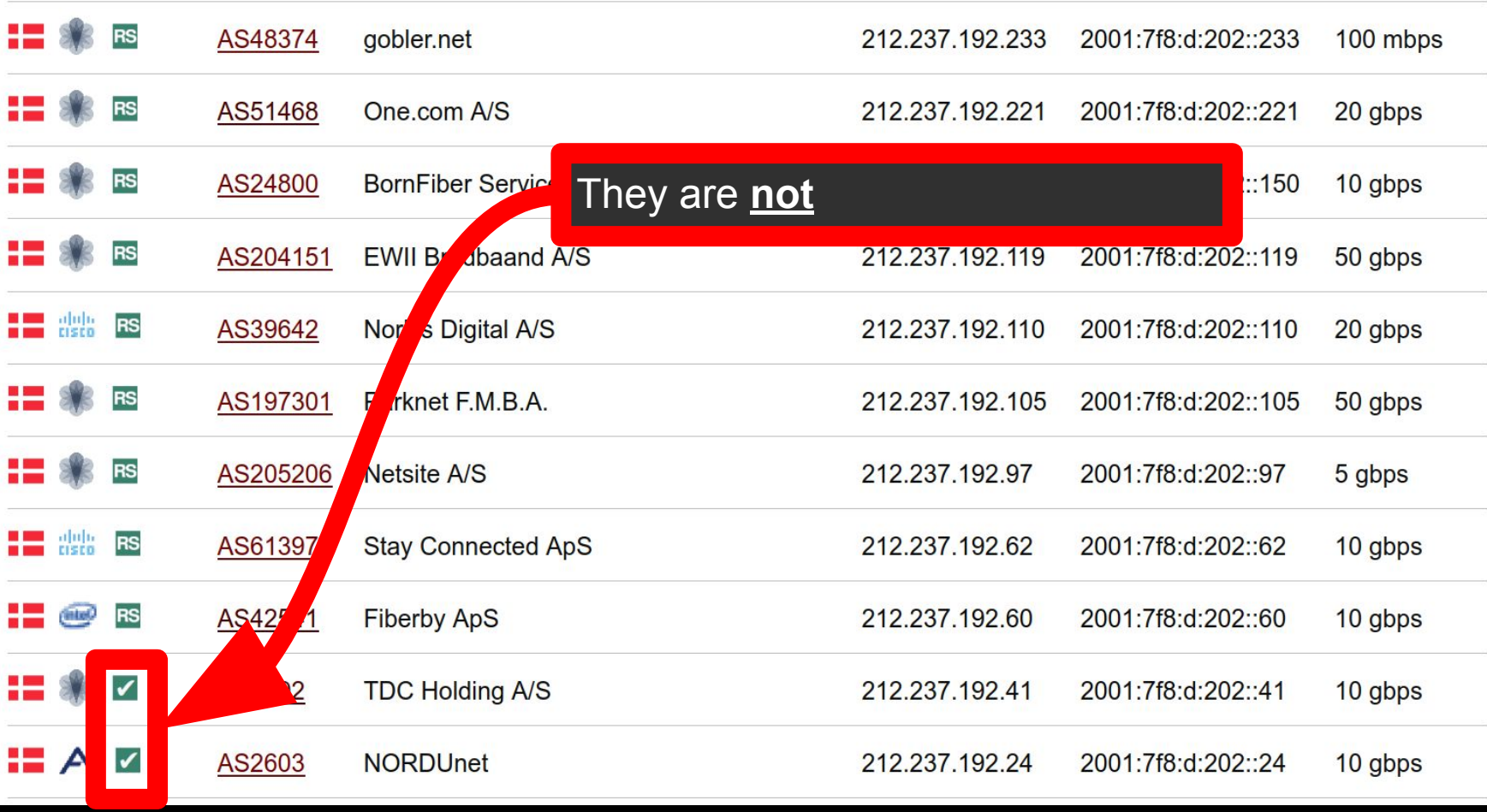
[1] https://bgp.tools/ixp/Netnod%20Copenhagen%20BLUE%20--%20MTU9K

9933dc1dafaf3a99bb5bddd73b039965

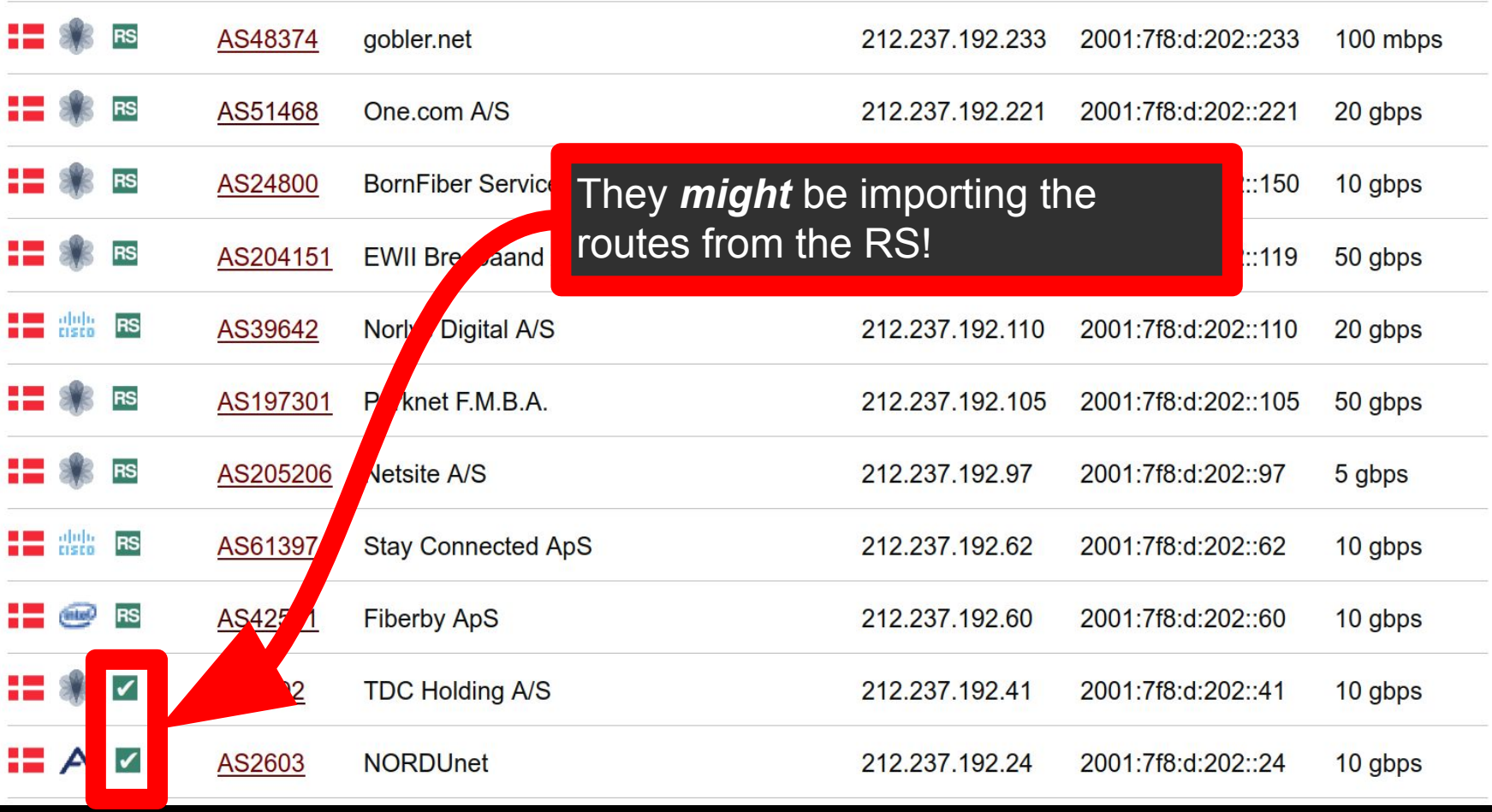


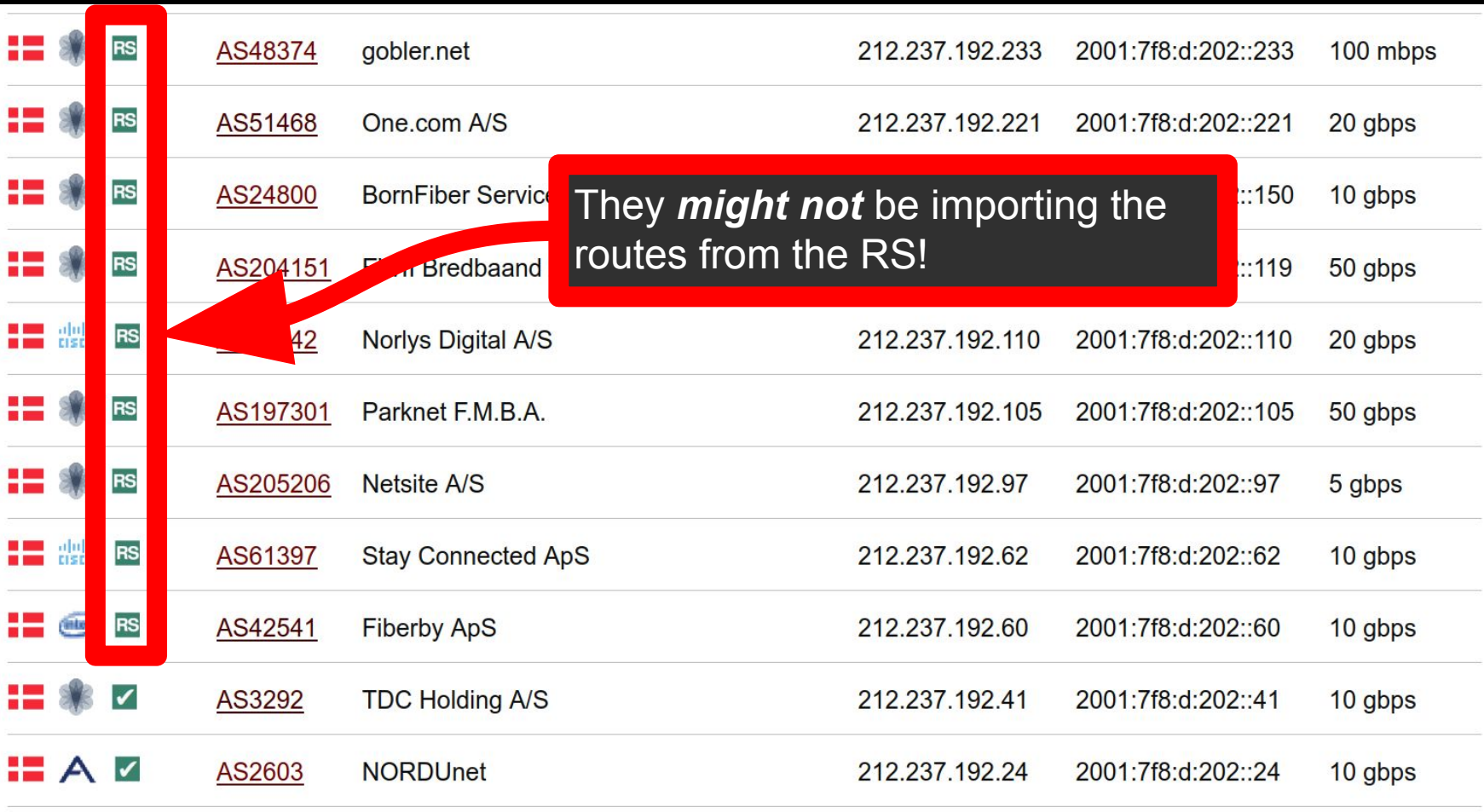

| | | | | | |
|---|---|---|---|---|---|
| RS | AS48374 | gobler.net | 212.237.192.233 | 2001:7f8:d:202::233 | 100 mbps |
| RS | AS51468 | One.com A/S | 212.237.192.221 | 2001:7f8:d:202::221 | 20 gbps |
| RS | AS24800 | BornFiber Servic | | ::150 | 10 gbps |
| RS | AS204151 | Bredbaand | | ::119 | 50 gbps |
| RS | 42 | Norlys Digital A/S | 212.237.192.110 | 2001:7f8:d:202::110 | 20 gbps |
| RS | AS197301 | Parknet F.M.B.A. | 212.237.192.105 | 2001:7f8:d:202::105 | 50 gbps |
| RS | AS205206 | Netsite A/S | 212.237.192.97 | 2001:7f8:d:202::97 | 5 gbps |
| RS | AS61397 | Stay Connected ApS | 212.237.192.62 | 2001:7f8:d:202::62 | 10 gbps |
| RS | AS42541 | Fiberby ApS | 212.237.192.60 | 2001:7f8:d:202::60 | 10 gbps |
| ✓ | AS3292 | TDC Holding A/S | 212.237.192.41 | 2001:7f8:d:202::41 | 10 gbps |
| ✓ | AS2603 | NORDUnet | 212.237.192.24 | 2001:7f8:d:202::24 | 10 gbps |

9933dc1dafaf3a99bb5bddd73b039965




9933dc1dafaf3a99bb5bddd73b039965


AS48374 gobler.net 212.237.192.233 2001:7f8:d:202::233 100 mbps
AS51468 One.com A/S 212.237.192.221 2001:7f8:d:202::221 20 gbps
AS24800 BornFiber Servic ::150 10 gbps
AS204151 EWII Bre aand ::119 50 gbps
AS39642 Norl Digital A/S 212.237.192.110 2001:7f8:d:202::110 20 gbps
AS197301 P knet F.M.B.A. 212.237.192.105 2001:7f8:d:202::105 50 gbps
AS205206 Netsite A/S 212.237.192.97 2001:7f8:d:202::97 5 gbps
AS61397 Stay Connected ApS 212.237.192.62 2001:7f8:d:202::62 10 gbps
AS425 1 Fiberby ApS 212.237.192.60 2001:7f8:d:202::60 10 gbps
2 TDC Holding A/S 212.237.192.41 2001:7f8:d:202::41 10 gbps
AS2603 NORDUnet 212.237.192.24 2001:7f8:d:202::24 10 gbps
They *might* be importing the routes from the RS!




9933dc1dafaf3a99bb5bddd73b039965

# How good is the reach for RS's for outbound heavy nets?

- This is the most simple thing to answer, because all you need to do is get all of the RIB's of all of the IXs RS's you are on (bgp.tools is currently on 113~ IXs), and perform a total unique route count vs the full table

- A full table as of the time of writing is
  - IPv4: 1,037,759 routes
  - IPv6: 236146 routes

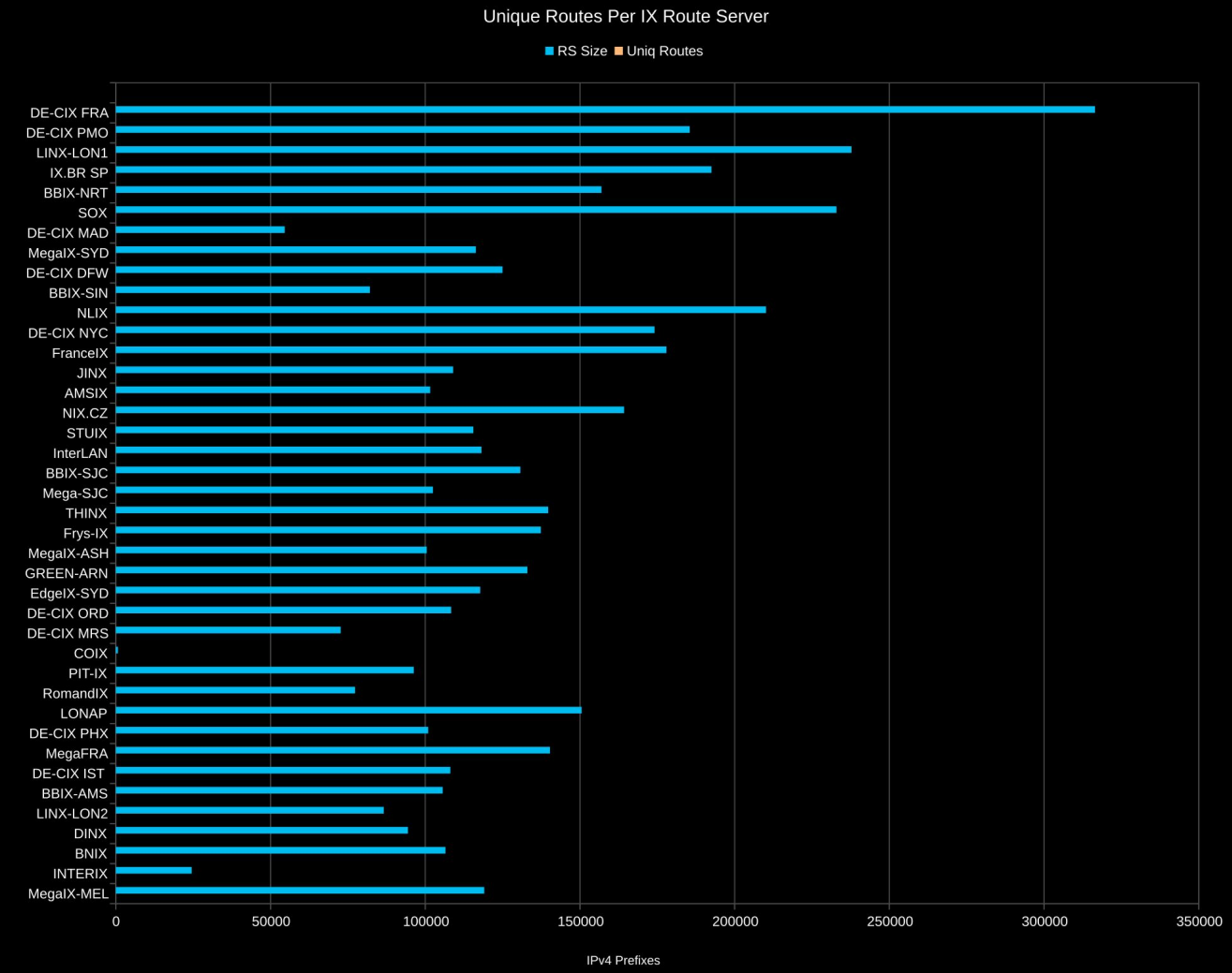
9933dc1dafaf3a99bb5bddd73b039965

# "Total" Outbound RS Reachability

- IPv4 - 56.6%
  - xxxx Reachable via RS
  - 1,037,759 Total Prefixes in DFZ

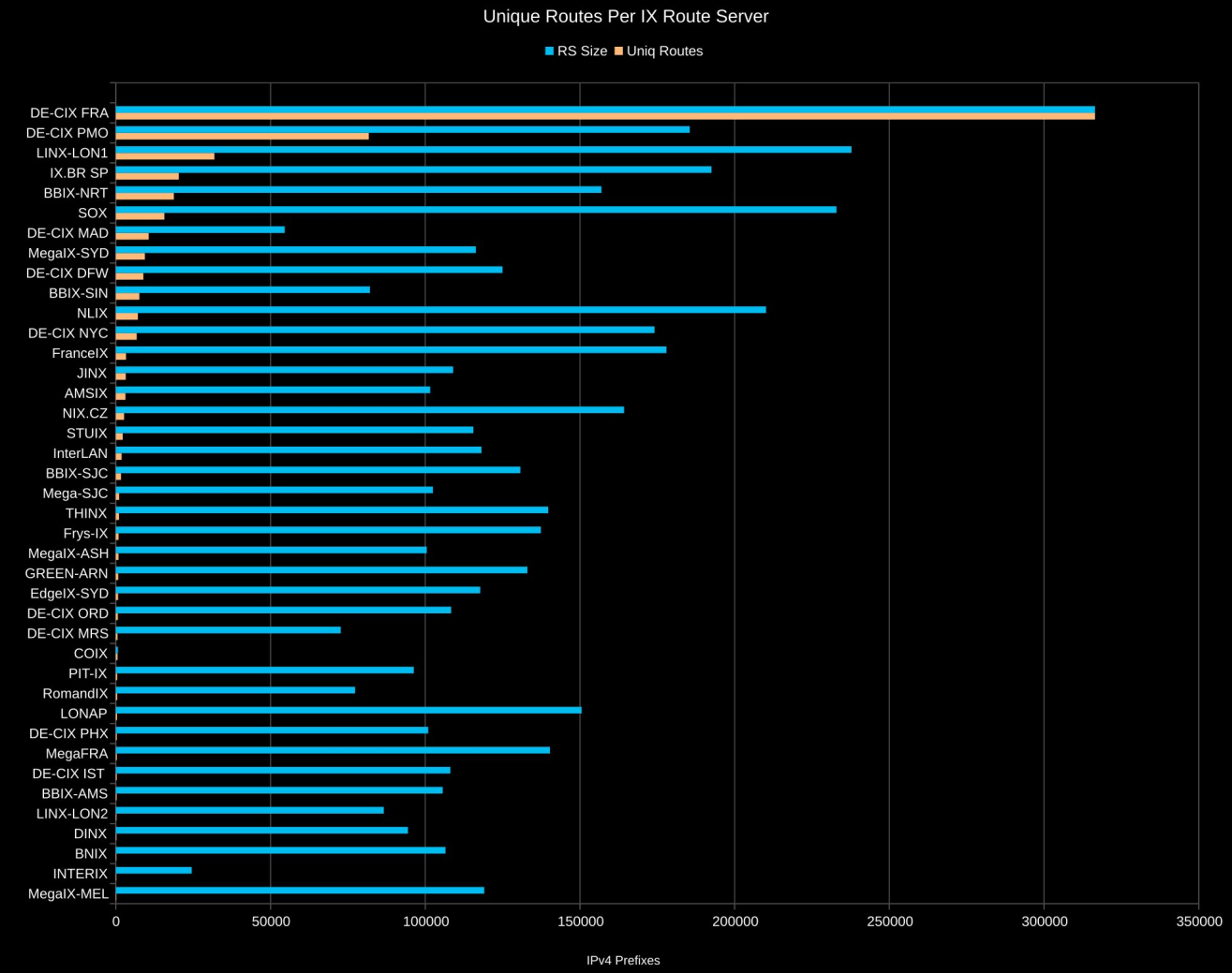
- IPv6 - 60%
  - xxxx Reachable via RS
  - 236,146 Total Prefixes in DFZ

9933dc1dafaf3a99bb5bddd73b039965




Most exchanges have route servers at least 100k IPv4 routes, Cool!

But that is overlooking a quirk…




If we instead look at what we gain by adding each exchange, the picture suddenly looks less great,

MegaIX-MEL has only 161 "novel" prefixes when all of the exchanges above it are taken into consideration!

The long tail here is very long

# Obvious (to me) cavates to this

1. This is a survey of 125 exchanges (a lot of exchanges)
    a. You are not likely to go out and build a network like this to offload this traffic, you would just buy more transit instead
    b. Also this 125 exchanges is missing a few notable groups of exchanges, notably, Equinix

2. Just because it's 60% of your prefixes does not mean it is 60% of your traffic
    a. As far as I can tell for eyeball traffic profiles, 50%+ of all traffic is concentrated in just 5 ASNs
        i. Like Meta, Akamai, Google, Netflix, Amazon | (The "Magna" networks?)

3. This calculation does not account for people pushing more specifics into the IX RS for traffic engineering, A technique very popular in markets like Brazil

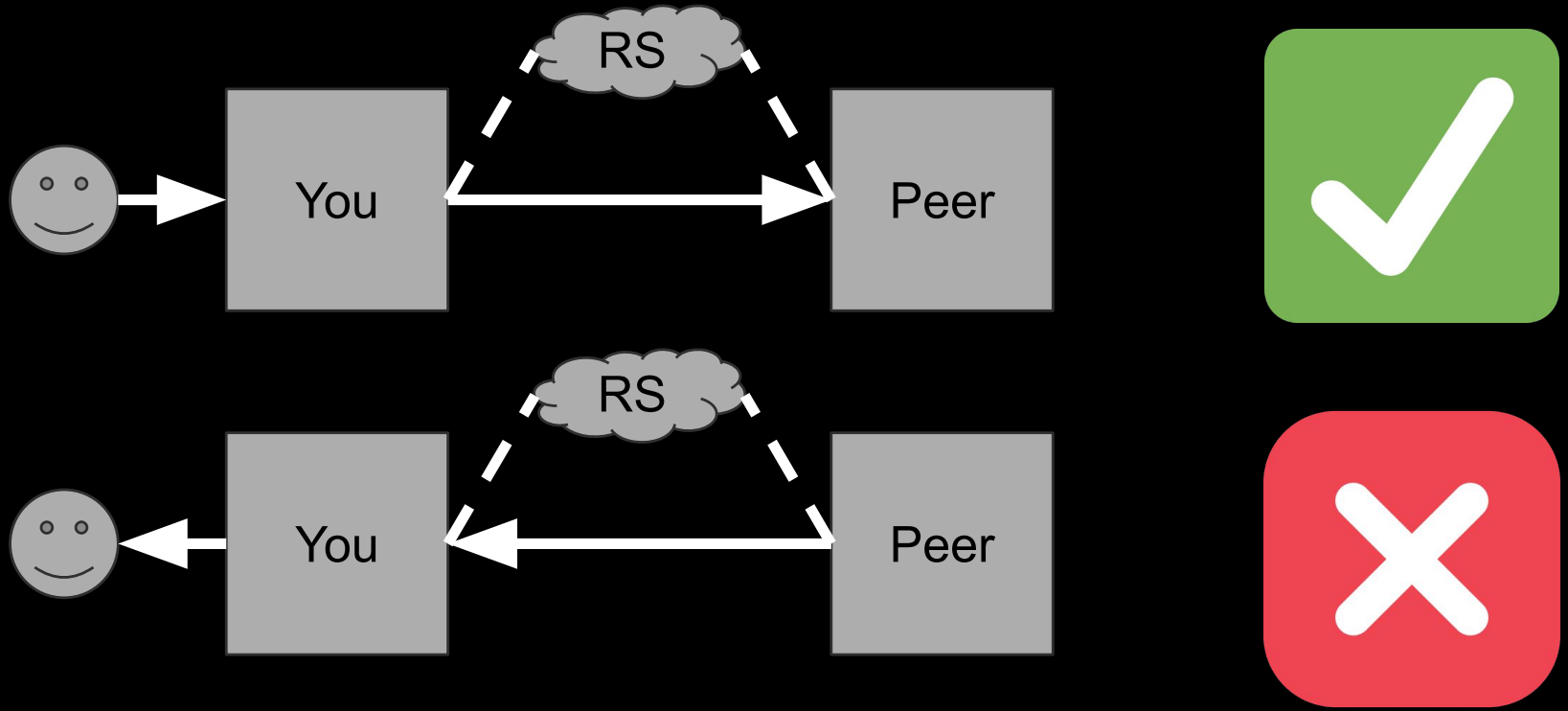
9933dc1dafaf3a99bb5bddd73b039965

# Full list of exchanges

ACT-IX (Canberra AU)
AMS-IX
B-IX
BBIX Amsterdam
BBIX Chicago
BBIX Dallas
BBIX London
BBIX Singapore
BBIX Tokyo
BBIX US-West
BCIX
BIX.BG
BNIX
BreizhIX
CANIX Montreal
CANIX Ottawa
CANIX Toronto
COIX
CINX
DE-CIX Barcelona
DE-CIX Chicago
DE-CIX Dallas
DE-CIX Dusseldorf
DE-CIX Frankfurt
DE-CIX Hamburg
DE-CIX Istanbul
DE-CIX Leipzig
DE-CIX Lisbon
DE-CIX Madrid
DE-CIX Marseille
DE-CIX Munich
DE-CIX New York
DE-CIX Palermo
DE-CIX Phoenix
DE-CIX Richmond
DINX
EdgeIX - Adelaide
EdgeIX - Auckland
EdgeIX - Brisbane
EdgeIX - Melbourne
EdgeIX - Perth
EdgeIX - Sydney
FD-IX - Indianapolis
FIXO
France-IX AURA
France-IX Lille
France-IX Marseille
France-IX Paris
France-IX Toulouse
Frys-IX
GetaFIX Cebu
GetaFIX Davao
GraniteIX
IndyIX
INTERIX
InterLAN-IX
IRAQ-IXP
IX.br São Paulo
IXP.mk
JINX
Lillix
LINX LON1
LINX LON2
LONAP
LU-CIX
MegaIX Adelaide
MegaIX Ashburn
MegaIX Atlanta
MegaIX Auckland
MegaIX Bay Area
MegaIX Berlin
MegaIX Brisbane
MegaIX Charlotte
MegaIX Chicago
MegaIX Dallas
MegaIX Denver
MegaIX Dusseldorf
MegaIX Frankfurt
MegaIX Hamburg
MegaIX Las Vegas
MegaIX Los Angeles
MegaIX Melbourne
MegaIX Miami
MegaIX Munich
MegaIX New York
MegaIX Perth
MegaIX Seattle
MegaIX Singapore
MegaIX Sofia
MegaIX Sydney
MegaIX Toronto
MSK-IX Moscow
Netnod Copenhagen
Netnod Gothenburg
Netnod Helsinki
Netnod Oslo
Netnod Stockholm
Netnod Sundsvall
NIX.CZ
NIX.SK
NL-ix
NMBINX
NSW-IX
ONIX
PatagoniaIX
PIT-IX
QLD-IX
NMBINX
NSW-IX
ONIX
PatagoniaIX
PIT-IX
QLD-IX
RomandIX
SA-IX
SIX.SK
SONIX Stockholm
SOX Serbia
STUIX
Stuttgart-IX
THINX Warsaw
TOP-IX
VIC-IX
WA-IX
YXEIXRomandIX
SA-IX
SIX.SK

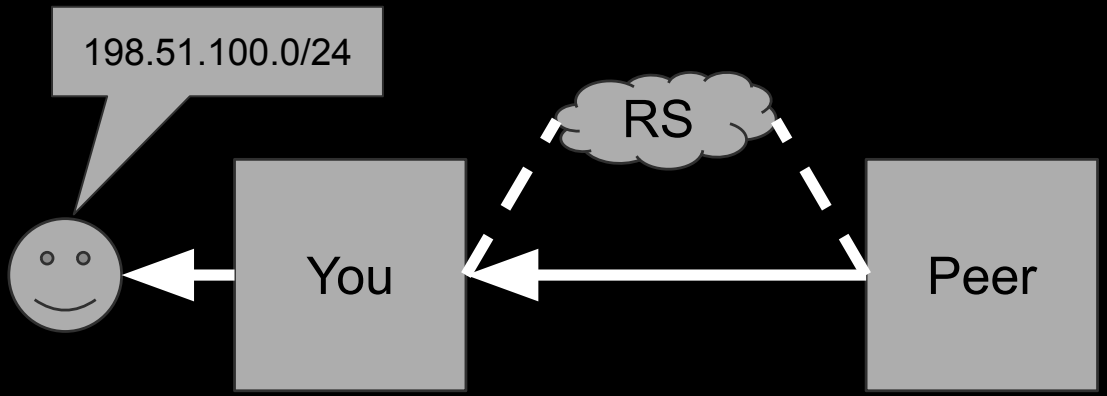
9933dc1dafaf3a99bb5bddd73b039965

# Traffic directions covered



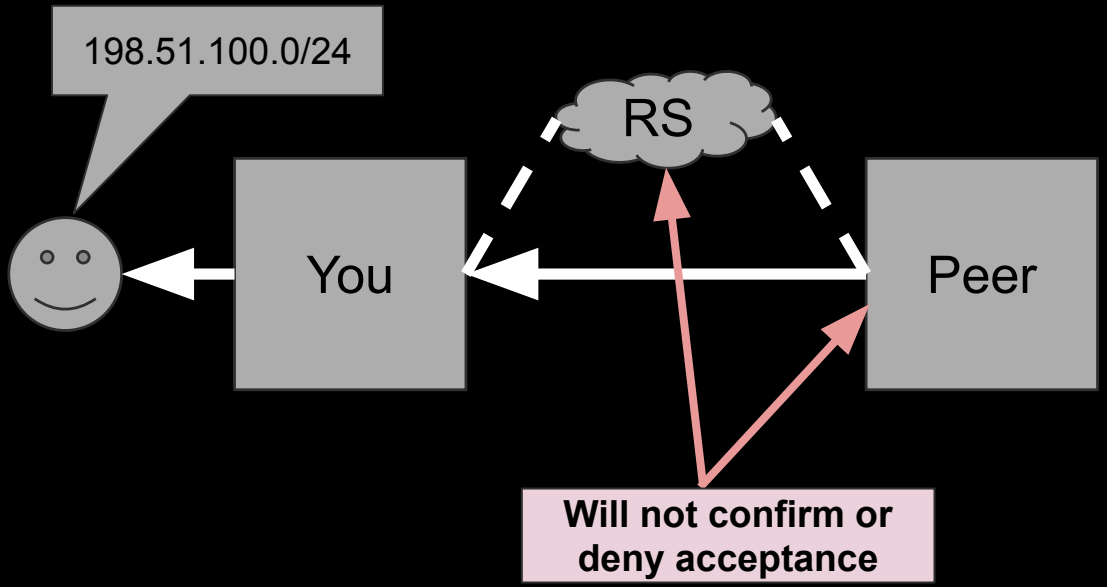

9933dc1dafaf3a99bb5bddd73b039965

# Inbound will need a different strategy

- 

# Inbound will need a different strategy

- BGP (famously) has no feedback signal to a peer for "yes I accepted a prefix"



# Inbound will need a different strategy

- BGP (famously) does not have a feedback signal to a peer for "yes I accepted a prefix"

- So we need to find another way to see if they have accepted the prefix
  - One way could be to use the bgp.tools live data set, however even though this is big (3k sessions), that is not big enough to cover the "whole" internet

- So we need to figure out a way to do test this on the data plane!

9933dc1dafaf3a99bb5bddd73b039965

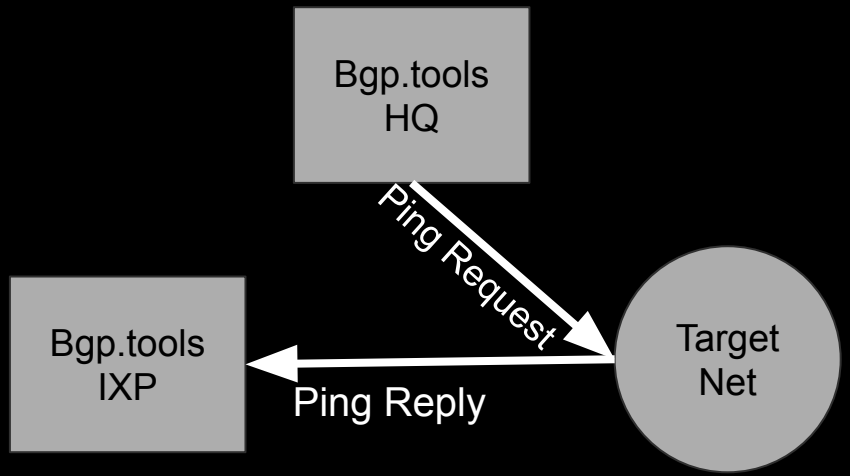
# Actual strategy

- Announce a prefix to all route servers
- From somewhere else, send/spoof ICMP pings to internet address, with the source IP being that RS only prefix
- If they are accepting the IX RS route, the reply will go back to the IX node, if they don't the reply has nowhere to go

9933dc1dafaf3a99bb5bddd73b039965

# Actual strategy

- Announce a prefix to all route servers
- From somewhere else, send/spoof ICMP pings to internet address, with the source IP being that RS only prefix
- If they are accepting the IX RS route, the reply will go back to the IX node, if they don't the reply has nowhere to go
-




9933dc1dafaf3a99bb5bddd73b039965

# Actual strategy

- Announce a prefix to all rout rs
- From somewhere else et address, with the source IP being that RS only
- If they are acceptin node, if they don't the reply as nowhere

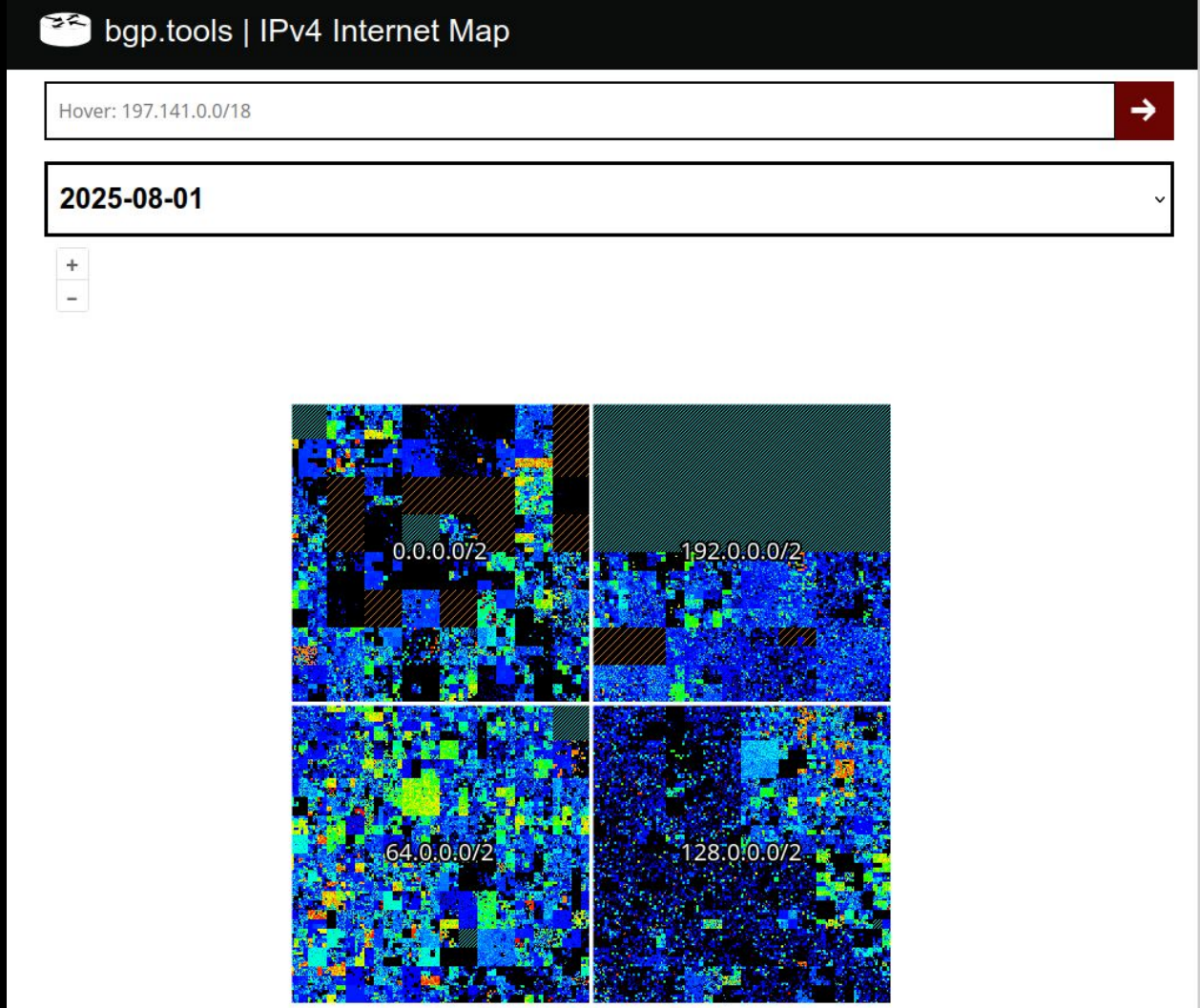
203.10.63.0/24

Thanks to Geoff Huston/APNIC!

# Optimisations

- To speed up things, we will only be pinging 1 IP out of each /24 (one that is known to reply)

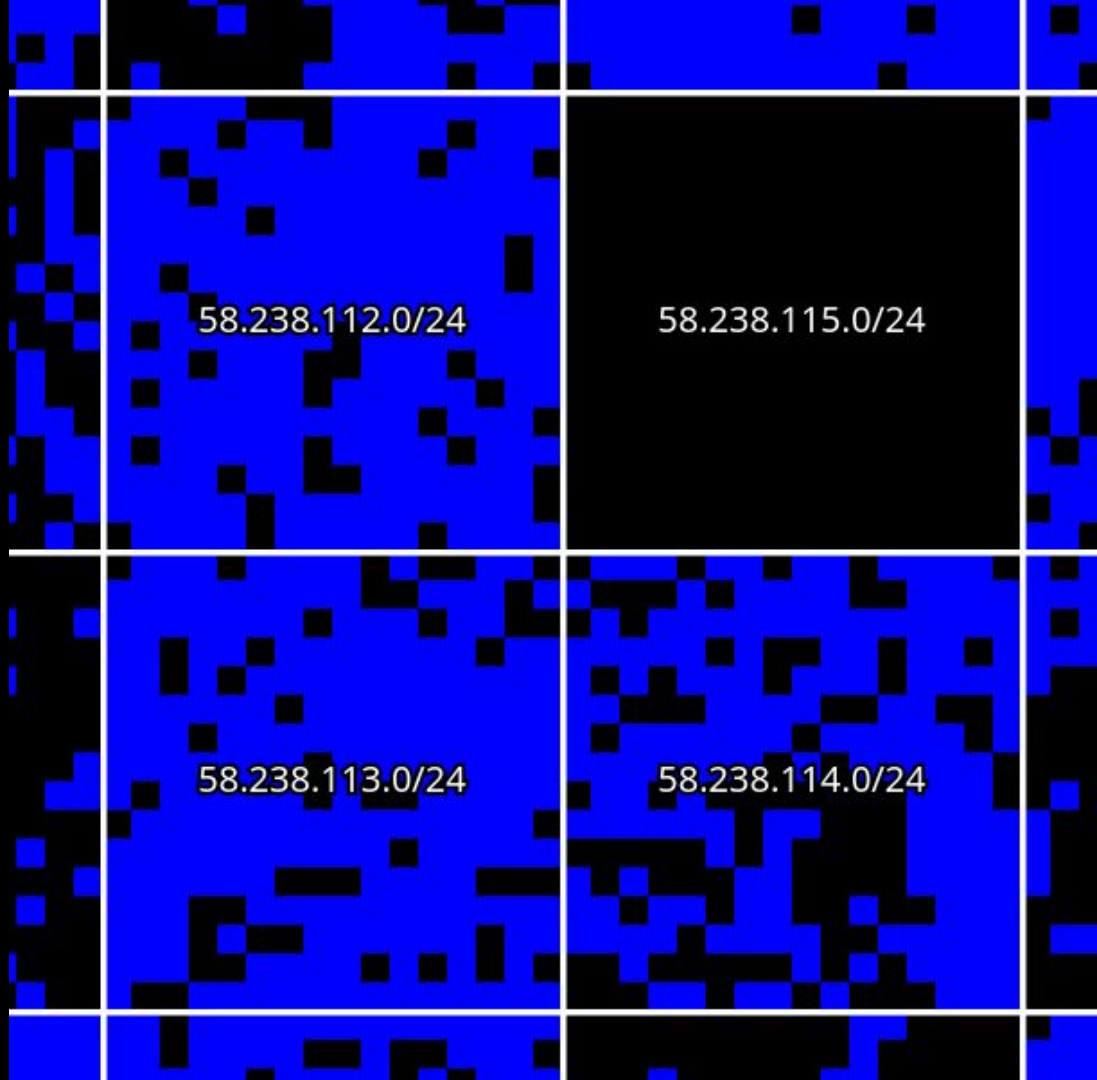
- Bgp.tools already knows what responds to pings because there is map.bgp.tools !



# Getting left behind

- Because not everything responds to ICMP, some prefixes are just left behind in this test because there is no easy way to test them

- This is sadly common for CGNAT blocks



# If you use all* RS-es, you get

- 95% (988,942) of IPv4 prefixes can be tested
- 14% (148,605) of prefixes can reach you if you are IX RS only
- 85% of prefixes can't use route servers paths
- Using 106 internet exchanges (where the RS's worked for me)

9933dc1dafaf3a99bb5bddd73b039965

# If you use all* RS-es, you get

- 95% (988,942) of IPv4 prefixes can be tested
- 14% (148,605) of prefixes can reach you if you are IX RS only
- 85% of prefixes can't use route servers paths
- Using 106 internet exchanges (where the RS's worked for me)

IPv6:

- 29% (68,890) of IPv6 prefixes can be tested
- 14% (34,451) of prefixes can reach you if you are IX RS only

Finding "pingable" addresses in IPv6 blocks is a hard problem to solve, the IPv6 number may be more "fluffy" than expected because of this, it could be that 50% of all v6 prefixes are reachable if we could find a address to ping them on…

9933dc1dafaf3a99bb5bddd73b039965

# Most packets go to the following exchanges:

| **Internet Exchange** | Prefixes Reply Count | Percent of all prefixes |
|---|---|---|
| BBIX Singapore | 18729 | 10.05% |
| IX.br (PTT.br) São Paulo | 18062 | 9.69% |
| BBIX Tokyo | 16931 | 9.09% |
| DE-CIX Frankfurt | 15085 | 8.10% |
| LINX LON1 | 11547 | 6.20% |
| MSK-IX Moscow | 11118 | 5.97% |
| AMS-IX | 8004 | 4.29% |
| NL-ix | 7922 | 4.25% |
| DE-CIX Dallas | 4949 | 2.65% |
| DE-CIX New York | 4383 | 2.35% |
| BIX.BG | 4071 | 2.18% |
| DE-CIX Madrid | 3108 | 1.66% |
| THINX Warsaw | 3033 | 1.62% |

9933dc1dafaf3a99bb5bddd73b039965

# Thoughts

- The difference between the EPF host IXs and all exchanges is 27,473~ routes
  - Now those 27,473 routes may matter a lot!
  - But clearly a lot of this is *dominated* by a small number of networks with large downstream customer counts who also import route servers
- Many people export to route servers, few people import from them
  - 566,000 unique prefixes available on Route Servers
  - 144,526 unique prefixes are behind networks that import from Route Servers
- That's a huge difference!
- Some of this is also limited by some prefixes (maybe the "juicy" CGNAT) ones being untestable

9933dc1dafaf3a99bb5bddd73b039965

# Bonus Thoughts

- Because this system uses the same pipeline bgp.tools uses to detect "naughty packets" on exchanges, I also know what networks these route server replies came from
- The results surprised me! On the All IXPs run the top source networks were:
  - BBIX Tokyo AS9498 / Bharti Airtel Ltd.
  - LINX LON1 AS58453 / China Mobile International
  - BBIX Singapore AS9002 / RETN Limited
  - BBIX Singapore AS18403 / FPT Telecom Company
  - DE-CIX Chicago AS6939 / Hurricane Electric LLC
  - BBIX US-West AS7713 / PT Telkom Indonesia Tbk
  - BBIX Tokyo AS3786 / LG DACOM Corporation
  - NL-ix AS4766 / Korea Telecom
  - Etc, In total 7185 different routers send replies

9933dc1dafaf3a99bb5bddd73b039965

# Thanks!

Questions? Comments? Stories?
Shy? Email ben@benjojo.co.uk
( or fedi/mastodon @benjojo@benjojo.co.uk )

9933dc1dafaf3a99bb5bddd73b039965